16.01.2016

Conjugate distributions

Example:

$$f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Beta prior: $f(p) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}$

The posterior distribution becomes also a beta distribution with updated parameters

$$f(p|x) = \frac{\Gamma(\tilde{\alpha}+\tilde{\beta})}{\Gamma(\tilde{\alpha})\Gamma(\tilde{\beta})} p^{\tilde{\alpha}-1} (1-p)^{\tilde{\beta}-1}$$

$\tilde{\alpha} = \alpha + x$

$\tilde{\beta} = \beta + n - x$

$\Rightarrow$ We say that the family of beta distributions is conjugate to the binomial distribution

Similarly:

$$(y_0, \dots, y_4) \,|\, (q_0, \dots q_4) \sim \text{Multinomial}(m, q_0, \dots q_4)$$

$$q \in (0,1) \quad \sum q_i = 1$$

Dirichlet prior $(q_0, \dots q_4) \sim \text{Dirichlet}(\alpha_0, \dots, \alpha_4)$

Posterior is also Dirichlet

$$(q_0, \dots, q_4) \,|\, (y_0, \dots y_4) \sim \text{Dirichlet}(\tilde{\alpha}_0, \dots \tilde{\alpha}_4)$$

$\tilde{\alpha}_i = \alpha_i + y_i$

The family of Dirichlet distributions is conjugate to the multinomial distribution.

Conditional conjugacy

Let $x_1, \dots, x_n \sim N(\mu, \sigma^2)$ indep.

Assume μ and σ^2 independent a priori, i.e

$$\pi(\mu, \sigma^2) = \pi(\mu) \cdot \pi(\sigma^2)$$

and assume

$\mu \sim N(\mu_0, \tau^2)$ (conjugate prior if σ^2 would be known)

$\sigma^2 \sim IG(\alpha, \beta)$ (conjugate prior if μ would be known)

$IG(\alpha, \beta)$: $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\exp(-\frac{\beta}{x})}{x^{\alpha+1}}$

$x \sim Ga(\alpha, \beta) \Rightarrow \frac{1}{x} \sim IG(\alpha, \beta)$

Then the posterior does not have a well known form.

$$\pi(\mu, \sigma^2 \mid x_1, \dots, x_n) \propto \frac{1}{\sigma^n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \cdot \exp\left(-\frac{1}{2\tau^2}(\mu - \mu_0)^2\right) \cdot \frac{\exp\left(-\frac{\beta}{\sigma^2}\right)}{(\sigma^2)^{\alpha+1}}$$

This cannot be written like a product of a normal for μ and an inverse gamma for σ^2.

However,

$$\pi(\underline{\mu} \mid x_1,\dots,x_n,\sigma^2) \propto \pi(\mu,\sigma^2,x_1,\dots,x_n)$$

$$\propto \exp\left(-\frac{1}{2\tau^2}(\mu-\mu_0)^2\right)\cdot \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

this is a normal distribution

mean over x

$$\mu \mid x_1,\dots,x_n,\sigma^2 \sim N\left(\frac{\frac{1}{\tau^2}\mu_0+\frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}},\ \frac{1}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}}\right)$$

Combining quadratic forms:

$$(x-a)^T A(x-a) + (x-b)^T B(x-b)$$

$$= (x-c)^T C(x-c) + (a-b)^T A C^{-1} B(a-b)$$

$$C = A+B$$

$$c = (A\cdot a + Bb)/C$$

$$\pi(\sigma^2 \mid x_1,\dots,x_n,\mu) \propto \pi(\mu,\sigma^2,x_1,\dots,x_n)$$

$$\propto \frac{\exp\left(-\frac{\beta}{\sigma^2}\right)}{(\sigma^2)^{\alpha+1}}\cdot\frac{1}{\sigma^n}\cdot\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$= \frac{\exp\left(-\frac{1}{\sigma^2}\left(\beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right)}{(\sigma^2)^{\alpha+\frac{n}{2}+1}}$$

$$\Rightarrow \sigma^2 \mid x_1\dots x_n,\mu \sim IG\left(\alpha+\frac{n}{2},\ \beta+\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

The distributions $\pi(\mu \mid x_1\dots x_n,\sigma^2)$ and $\pi(\sigma^2 \mid x_1\dots x_n,\mu)$ are called the full conditionals and we will later see how ~~to use these~~ it is useful that these are explicitly known.

Noninformative prior distributions

$\theta \sim \text{Unif}$ (which is possibly improper)

Let $\varphi = h(\theta)$

$$f(\varphi) = f(h^{-1}(\varphi)) \cdot \underbrace{\left| \frac{d h^{-1}(\varphi)}{d\varphi} \right|}_{\text{only constant for linear function } h}$$

$\Rightarrow$ That means for a non-linear function $f(\varphi)$ will not be uniform anymore

$\hat{=}$ a bit strange